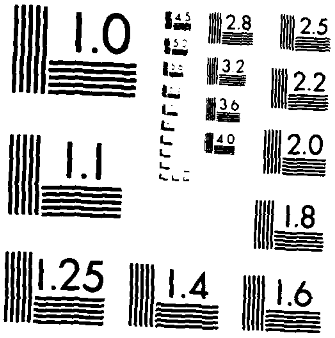
AD-A129 938 TRYING FOR SPEAKER INDEPENDENCE IN THE USE OF SPEAKER DEPENDENT VOICE REC..(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA G K POOCK ET AL. DEC 82 NPS55-82-032 1/1

UNCLASSIFIED MIPR-TB-024 F/G 17/2 NL


1.0
4.5
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

NPS55-82-032

NAVAL POSTGRADUATE SCHOOL

Monterey, California

TRYING FOR SPEAKER INDEPENDENCE IN THE USE OF

SPEAKER DEPENDENT VOICE RECOGNITION EQUIPMENT

by

Gary K. Poock
Norman D. Schwalm
B. Jay Martin
Ellen F. Roland

December 1982

Approved for public release; distribution unlimited.

Prepared for:

9th Infantry Division
Fort Lewis, WA 98433

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral J. J. Ekelund
Superintendent

D. A. Schrady
Provost

This work was performed by the authors at the Naval Postgraduate School, Monterey, California.

Reproduction of all or part of this report is authorized.

Gary Poock

Gary K. Poock, Professor
Department of Operations Research

Norman D. Schwalm
Perceptronics

B. Jay Martin

B. Jay Martin
Perceptronics

Ellen F. Roland

Ellen F. Roland
Rolands and Associates

Reviewed by:

Kneale T. Marshall, Chairman
Department of Operations Research

Released by:

William M. Tolles

William M. Tolles
Dean of Research

Ellen F. Roland performed part of this work under NPS Contract N00228-82-C-6418 entitled "Research and development study of the feasibility of using computer voice entry".

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER
NPS55-82-032

2. GOVT ACCESSION NO.
[illegible]

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
TRYING FOR SPEAKER INDEPENDENCE IN THE USE OF SPEAKER DEPENDENT VOICE RECOGNITION EQUIPMENT

5. TYPE OF REPORT & PERIOD COVERED
Technical

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Gary K. Poock
Norman D. Schwalm
B. Jay Martin
Ellen F. Roland

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Naval Postgraduate School
Monterey, CA 93940

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
MIPR TB-024

11. CONTROLLING OFFICE NAME AND ADDRESS
9th Infantry Division
Fort Lewis, WA 98433

12. REPORT DATE
December 1982

13. NUMBER OF PAGES

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
VTAG
VOICE RECOGNITION/INPUT
TACFIRE
SPEAKER DEPENDENCE
SPEAKER INDEPENDENCE

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
This report discusses the results of an experiment to determine the possibilities of obtaining some speaker independence using speaker dependent voice recognition equipment. The results revealed about 99% accuracy when the user's speech templates were in memory along with those of four other users. If the user's voice patterns were not in memory but those of the four other users still were in memory, recognition accuracy still hovered around 95%.

FOREWORD

This is one of several reports looking at the feasibility of using current state-of-the-art voice recognition technology as a possible method for data entry into the Army's TACFIRE system, but it is also applicable to other similar systems as well.

If voice recognition equipment were installed and used for voice data entry at the artillery control console in the TACFIRE van, the question was asked if it would be possible for multiple users to then use the system for voice data entry. To enable multiple users to use the same voice recognition machine, one would really want a speaker independent system which allows any user to speak to it. However, such systems with vocabularies of several hundred utterances are not commercially available. Some speaker independent systems with small vocabularies of ten to twenty utterances are available, but would not satisfy the needs in this case, and they are also very expensive.

Therefore, this report examines the possibility of using commercially available speaker dependent systems which have relatively high recognition accuracy for vocabularies of a few hundred utterances, and which are reasonably priced for a few thousand dollars.

The question examined was how well a speaker dependent system would work for multiple users when used as a speaker independent system.

TABLE OF CONTENTS

Page

EXECUTIVE SUMMARY iii

1. INTRODUCTION 1-1

1.1 Background 1-1
1.2 Problem 1-2
1.3 Objectives 1-2

2. METHOD 2-1

2.1 Subjects 2-1
2.2 Apparatus 2-1
2.3 Experimental Design 2-1
2.4 Procedure 2-2

2.4.1 Training 2-2
2.4.2 Testing 2-4

2.5 Independent and Dependent Variables 2-4

3. RESULTS 3-1

3.1 Overview 3-1
3.2 Total Errors 3-2
3.3 Nonrecognitions (Rejections) 3-7
3.4 Misrecognitions 3-11

4. DISCUSSION 4-1

4.1 Total Errors 4-1
4.2 Nonrecognitions 4-2
4.3 Misrecognitions 4-2

5. CONCLUSIONS 5-1

6. REFERENCES 6-1

APPENDIX A -- LIST OF UTTERANCES

EXECUTIVE SUMMARY

The purpose of the present study was to determine the accuracy rate of current voice recognition (henceforth, VR) systems if the speaker's input was compared to a group of speech patterns including (1) only the patterns of that speaker (entered during training), (2) the patterns of that speaker as well as 4 other speakers who had trained the same utterances, and (3) only the patterns of 4 other speakers (i.e., excluding the speaker's own pattern). The last condition is defined for purposes of this study as the speaker independence mode. It is conceivable that future uses of VR equipment may include command, control, and communication (C^3) centers, where it may be impractical to retain separate speech files for all users.

The findings suggest that nonrecognitions (e.g., errors where the system rejects the input and says, in effect, "I don't understand you, say it again") were not affected by comparing the speaker's input to a group of patterns including his own, and increased less than 4% when comparing the speaker's input to a group of patterns not including his own.

Misrecognitions (i.e., errors where the system accepts the input but mistakes it for a different input) remained near or below 1% and were not significantly affected by any of the comparisons.

It was concluded that current VR equipment may be used with about 99% accuracy in situations where it is impractical to access separate speech files for individual users since the speech patterns of all users are in the same file. Furthermore, current VR equipment may be used with 95% accuracy in speaker independent situations where the voice recognition device (henceforth, VRD) has no access to the current user's speech patterns. In this, the speaker independent mode, the more problematic error of misrecognitions is still held to a rate of only 1%.

These findings imply a great potential in the flexibility of VRD's and expansion in the number of users to whose speech the VRD can respond accurately. The results of the present study are based on data from subjects who underwent a training session, in which they may have become practiced at speaking to the VRD. This, in turn, could have optimized the VRD's recognition accuracy, Future research should investigate the accuracy rate of the speaker independent mode where a completely naive user tests the system (i.e., a user who has not trained the system). This would also allow researchers to determine the effects, if any, of the initial training session on accuracy of recognition.

1. INTRODUCTION

1.1 Background

In recent years, voice technology has developed to the extent that basic systems have now been used successfully in several industrial and military applications. With constant improvements being made in the capabilities of voice recognition systems, their use in a wider variety of settings is already being contemplated.

As the variety of settings widens, the requirements for the VRD become more diversified. One situation may require a VRD to recognize the speech of only one user who has throughly "trained" the system. Another situation might require the VRD to recognize the speech of several users, and, in some instances, to recognize the speech of a user for whom the VRD has no speech patterns recorded. In these cases it would be desirable for the VRD to be capable of recognizing the speech of as many users as possible, without an increase in errors due to the variance of speech patterns from user to user.

In another setting, perhaps for security purposes, a VRD might be required to recognize and respond to only a particular person or set of persons' speech. In this case it would be desirable for the VRD to recognize only the speech pattern(s) of the user(s) for whom it has patterns stored.

In any case, decisions must be made concerning the variety of stored speech patterns necessary for recognition of a user's speech in particular settings.

1.2 Problem

One way of optimizing accurate recognition of a particular user's speech might be to compare that user's inputs not only to his own speech patterns, but to the speech patterns of other users as well. Under some circumstances it is possible that a user's speech input might match the speech pattern of another user rather than his own. The circumstances that could lead to changes in one's speech patterns are common; slight changes in pronunciation, mood changes, and having a cold, are a few examples. On the other hand, comparing one user's speech to several users' speech patterns may have drawbacks. It is possible that with an increased number of user's patterns to compare to, the probability of misrecognizing an input (that would have otherwise simply been rejected) may increase. Misrecognitions are errors in which the VRD "thinks" it heard an utterance that matches one in memory, when, in fact, some other utterance was input. Misrecognitions are probably more problematic than nonrecognitions, in which the VRD simply rejects an input as unrecognizable. In the latter case no action is taken, whereas in the former case some inappropriate action may result.

The purpose of the current research was to explore various strategies for speech pattern comparisons, and the number and type of errors associated with each type of comparison.

1.3 Objectives

The specific objective of the present research was to assess empirically the accuracy with which currently available VRDs could interpret utterances when compared to: (1) the current speaker's patterns only; (2) the current speaker's patterns plus four other speaker's patterns; and (3) four other speaker's patterns only.

2. METHOD

2.1 Subjects

Fifteen volunteers (all males) were recruited from curriculums at the Naval Postgraduate School in Monterey, California.

2.2 Apparatus

A Threshold Technology model T600 voice recognition device was used in this study. The device was capable of storing 256 voice utterances of up to 2 seconds each. Fifty utterances were used in the present investigation. These utterances appear in Appendix A.

A Shure model SM10 "boom" microphone (mounted on a headset) was used as the input device. This microphone is supplied as standard equipment with the T600.

The Threshold system was linked to an IBM 3033 computer via a modem, allowing the experimenter to manipulate which set of speech patterns the Threshold would access when attempting to recognize the 50 utterances.

2.3 Experimental Design

A 3x3x6 mixed design, with repeated measures on two factors and replication on a third factor, was employed in this experiment. Test condition was a three-level within group variable. In the first test condition (S=Self Only) the VRD had access only to the speech patterns of the subject who was currently making voice inputs. In the second test condition (S+0 = Self plus Others) the VRD had access to the speech patterns of the current subject plus those of the other four members of his group. In the third test condition (0=Others Only) the VRD had access only to the speech patterns of the other four members of the current subject's group. Each subject performed 6 trials under each of the three test conditions, making trials the second within group variable with 6 levels. Three separate groups, with five subjects nested in each, were subjected to 6 trials under each of the three test conditions, making groups the between variable. Essentially, this resulted in multiple replications of the 3x6 portion of the experimental design. A summary of the experimental design appears in Figure 2-1.

2.4 Procedure

2.4.1 Training. The term "training," as used in discussions of voice recognition studies, refers to the process by which the speaker makes known to the recognizer the characteristics of his particular speech patterns for all the utterances he will be using. For the T600, this training procedure consists of entering 10 passes of each utterance (10x50 or 500 utterances for each subject) into the voice recognizer. The recognizer automatically enters these utterances into its "memory," and matches any subsequent utterances of the same vocabulary (in testing) with those in memory. Ideally, these subsequent utterances are matched with those in memory and the result is a correct response output on a CRT. In cases where the recognizer can not make this match, a nonrecognition or rejection occurs, and this results in a "beep" from the recognizer; in effect, the machine is saying "I don't understand that utterance--please say it again." Occasionally, however, the recognizer "thinks" it has matched an utterance with one in memory, but the match is incorrect. In this case, an incorrect response is output on the CRT, constituting what is known as a "misrecognition." Thus, two types of errors are possible: nonrecognitions (or rejections) and misrecognitions (or misinterpretations) of an utterance.

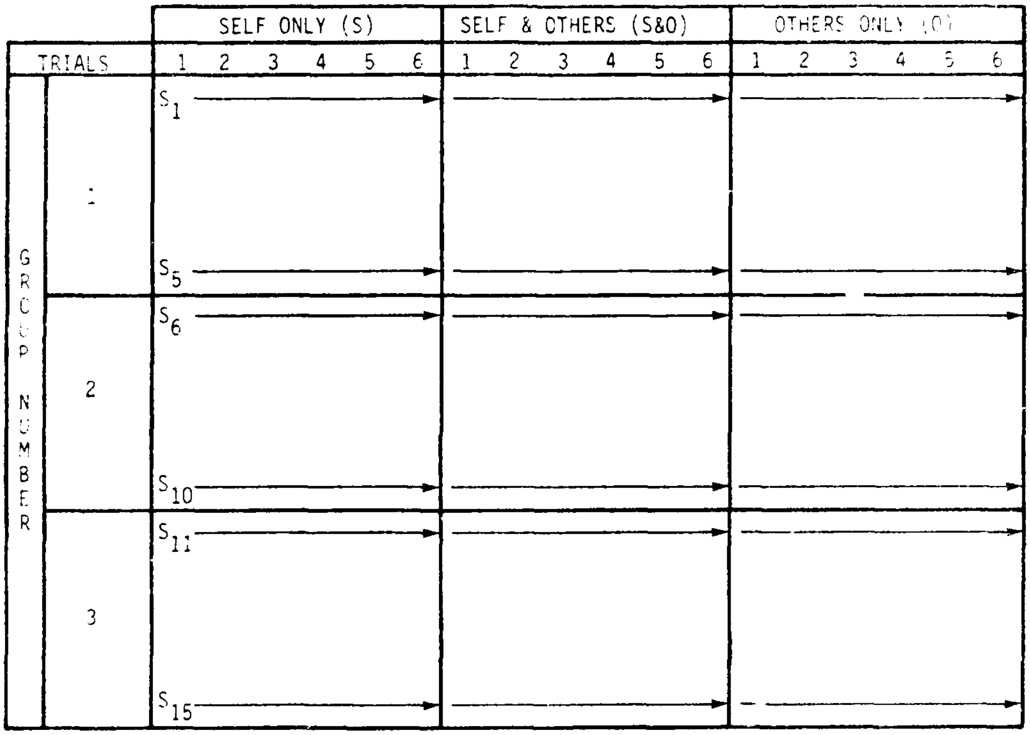


FIGURE 2-1
SUMMARY OF EXPERIMENTAL DESIGN

For training, each subject spoke 10 passes of each of 50 utterances into the VRD(total = 500 utterances per subject). This procedure took approximately one-half hour for each subject. Approximately ½ the subjects trained the system on Monday, and the other half on Tuesday.

Immediately after training, subjects made at least two passes of the entire 50 word vocabulary with the T600 memory open to only their own speech patterns (essentially a test session) to identify any problems in training of any particular utterance. Where the system produced correct responses on those two passes, the utterance was considered adequately trained. If errors occurred (of either type) a third pass was made. If less than two of three passes of any utterance was correct, that utterance was retrained.

2.4.2 Testing. After training, subjects tested the system. Each subject was scheduled to make two passes through the entire vocabulary list under each of the three test conditions on each of three successive days. These testing sessions were administered on Wednesday, Thursday, and Friday of the same week in which training took place. Thus, a total of six testing trials were run for each subject under each test condition. In this way, subjects were able to complete training and testing within one week.

2.5 Independent and Dependent Variables

The independent variables in this study were group, trials, and test condition: Self only, where the subjects tested the system with access only to their own speech pattern; Self + Others, where subjects tested the system with access to the speech patterns of their entire group (including their own); and Other Only, where subjects tested the system with access to the speech patterns of only the other members of their group.

The dependent variables in this study were nonrecognitions (or rejections). misrecognitions, and total errors, which was a linear combination of nonrecognitions and misrecognitions.

3. RESULTS

3.1 Overview

This section describes the results of the present study. All analyses were performed using the SPSS (Nie, Hull, Jenkins, Steinbrenner and Bent, 1975) and BMDP (Brown, Engelman, Frane, Hill, Jennrich and Toporek, 1981) statistical packages. All repeated measures analyses of variance procedures were performed using the arcsin transformation of raw data to stabilize the variance of the error terms (Neter and Wasserman, 1974). The mean error rates that appear in the figures, however, are untransformed. All a posteriori tests for significance between pairs of means were performed using the Scheffé procedures described in Bruning and Kintz (1977).

As defined earlier, nonrecognitions and misrecognitions by the voice recognition system may have distinctly different implications in an applied setting. To take an extreme example, in a weapons deployment activity, it would be far more desirable for the system to respond to an input error by nonrecognition (a "beep"), where the speaker is essentially told that he should repeat the input (or correct it), than for the system to misinterpret the input and to carry out some incorrect (and perhaps critical) command in error. Thus, it was considered essential to determine the effects of the independent variables on nonrecognitions and misrecognitions separately, as well as on total number of errors (nonrecognitions + misrecognitions).

Section 3.2 presents the data for total number of errors. Section 3.3 presents the results of analyses done on nonrecognitions or rejections, while Section 3.4 presents the results of analyses done on misrecognitions.

3.2 Total Errors

Table 3-1 presents the ANOVA summary table for total errors (nonrecognitions + misrecognitions). Significant main effects of test condition (F = 8.11, P < .002) and trials (F = 2.83, P < .05) are evident. No significant main effects for groups was found, but the groups by trials interaction was significant (F = 2.40, P < .05). Mean error rates (in percent) are shown in Table 3-2, and the main effects of test condition and of trials are portrayed graphically in Figures 3-1 and 3-2, respectively.

With regard to the main effect of test condition, a Scheffé test for significance between pairs of means was performed to determine between which pairs of means the significant differences lie. The results of this test indicated that significant differences existed between the self + others and the others only conditions, and between the self only and the others only conditions. The differences between the self only and the self + others conditions were not significant. Figure 3-1 portrays the relationship between the condition means. The figure shows that a significantly greater number of errors were recorded when subjects tested the VRD without access to their own speech patterns (i.e., speaker independent mode). Note, however, that the accuracy rate corresponding to that error rate still exceeds 95 percent.

Although the ANOVA indicated a significant trials effect, review of Figure 3-2 reveals no apparent systematic change over trials. A Scheffé test for significance between pairs of means detected no significant differences between any two trials. Evidently, the ANOVA is sensitive to the spurious nature of errors across trials. However, the difference between even the highest and lowest error rates over trials is not large enough to reach statistical significance in the post hoc Scheffé test. For further discussion on post hoc range tests, and lack of significance in post hoc tests where significance was reached in an analysis of variance, see J.L. Myers,

TABLE 3-1

ANOVA SUMMARY TABLE FOR TOTAL ERRORS

Source	df	MS	F
Group (G)	2	.49562	.79
Error	12	.62838	
Test Condition (C)	2	2.02629	8.11**
C x G	4	.13845	.55
Error	24	.24970	
Trials	5	.06445	2.83*
T x G	10	.05471	2.4*
Error	60	.02277	
C x T	10	.03729	1.64
C x T x G	20	.02973	1.31
Error	120	.02271	

* p < .05

** p < .01

TABLE 3-2.

MEAN TOTAL ERRORS (IN PERCENT) FOR TEST CONDITIONS BY TRIALS

	Self Only	Self + Others	Others Only	$\bar{x}$ Trials
Trial 1	00.667	01.333	05.067	02.353
2	01.067	01.067	03.600	01.911
3	01.200	00.667	05.600	02.489
4	01.200	00.400	05.333	02.311
5	00.667	00.133	04.267	01.689
6	00.800	00.800	05.200	02.267
x Test Conditions	00.934	00.733	04.845	Grand $\bar{x}$ 02.170

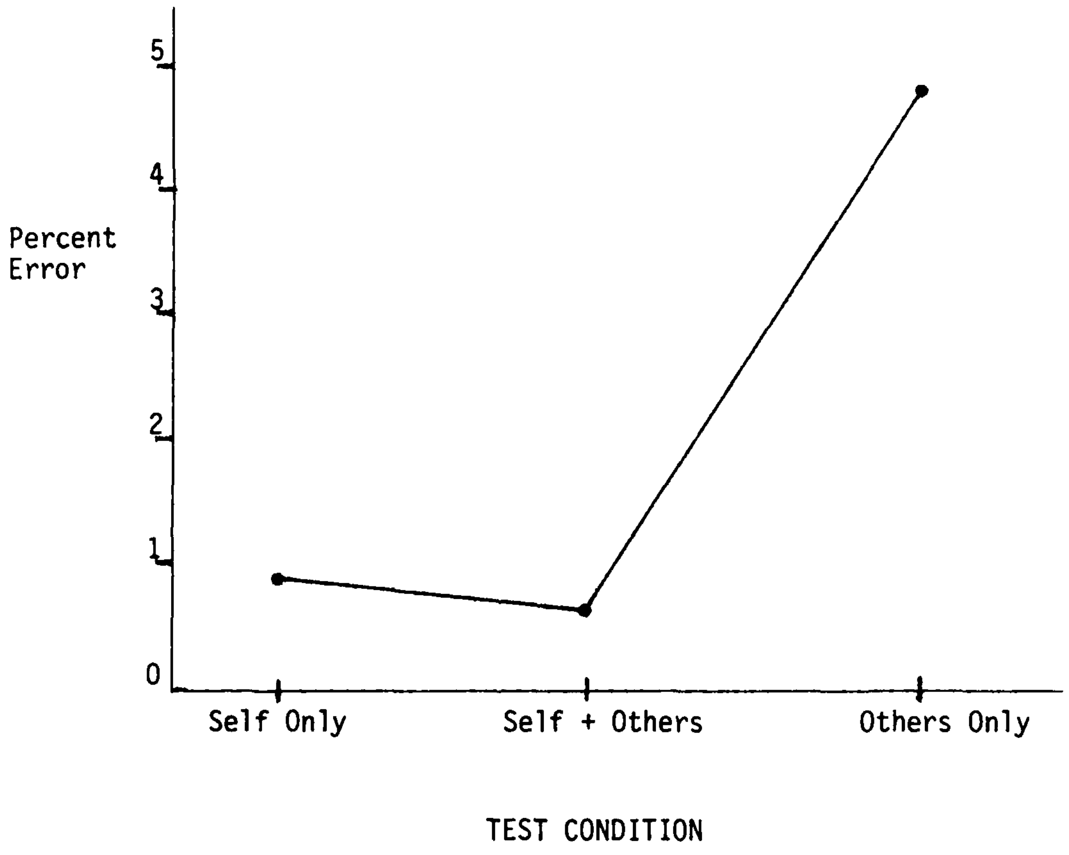


FIGURE 3-1.
TOTAL ERRORS BY TEST CONDITION.

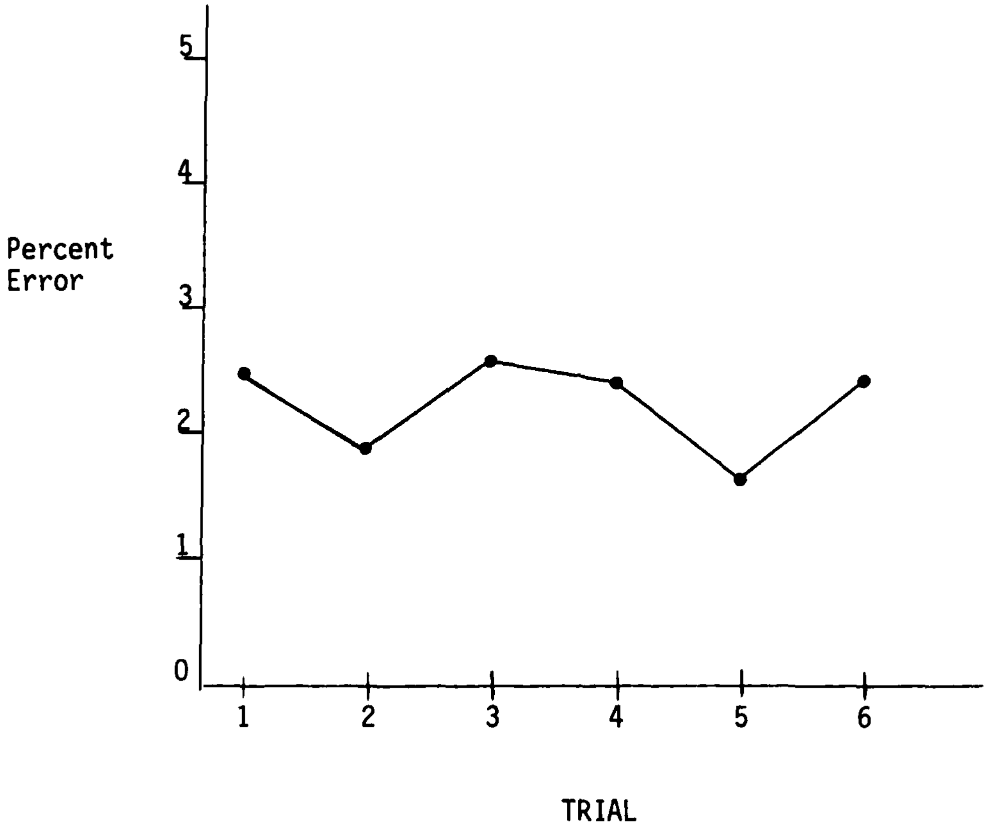


FIGURE 3-2.
TOTAL ERRORS BY TRIALS

1972. The authors considered the possibility that the Scheffe´ test may have been overly conservative, and subsequently performed other, less conservative tests. However, none of the appropriate range tests (e.g., Tuky, Newman-Keuls) revealed significant differences at the .05 level.

The groups by trials interaction also reached significance in the ANOVA (Table 3-1). Again, there were no interpretable or systematic effects, and the authors attach no practical significance to either the trials effect or the groups by trials interaction.

3.3 Nonrecognitions (Rejections)

An analysis of variance was performed on the nonrecognitions alone to determine the effects, if any, of the groups, trials, and test conditions. Table 3-3 presents the analysis of variance summary table for nonrecognitions.

A significant main effect of test condition ($F = 8.67$, $P < .01$) was found, but there were no significant main effects of groups or trials. Mean nonrecognition rates (in percent) are presented in Table 3-4, and the main effect of test condition is portrayed graphically in Figure 3-3.

With regard to the main effect of test condition, a Scheffe´ test for significance between pairs of means was performed to determine between which pairs of means the significant differences lie. The results of this test indicated that significant differences existed between the others only condition and the self plus others condition; and between the others only condition and the self only condition. The difference between the self only condition and the self plus others condition was not significant.

Review of Figure 3-3 indicates that nonrecognitions were reduced slightly when the system had access to the speech patterns of the entire group rather than just those of the current speaker. However, when the current speaker's

TABLE 3-3

ANOVA SUMMARY TABLE FOR NONRECOGNITIONS

SOURCE	df	MS	F
Group (G)	2	.08825	.31
Error	12	.28573	
Test Condition (C)	2	1.42606	8.67**
C x G	4	.02577	.16
Error	24	.16450	
Trials (T)	5	.02772	2.02
T x G	10	.02633	1.92
Error	60	.01370	
C x T	10	.01453	1.37
C x T x G	20	.01315	1.24
Error	120	.01057	

** $p < .01$

TABLE 3-4

MEAN NONRECOGNITIONS (IN PERCENT) FOR TEST CONDITIONS BY TRIALS

	Self Only	Self + Others	Others Only	$\bar{x}$ Trials
Trial 1	00.133	00.267	03.867	01.423
2	00.533	00.267	02.533	01.111
3	00.667	00.267	04.133	01.689
4	00.667	00.267	04.267	01.737
5	00.400	00.000	03.467	01.289
6	00.267	00.400	04.667	01.778
$\bar{x}$ Test Conditions	00.445	00.245	03.822	Grand $\bar{x}$ 01.504

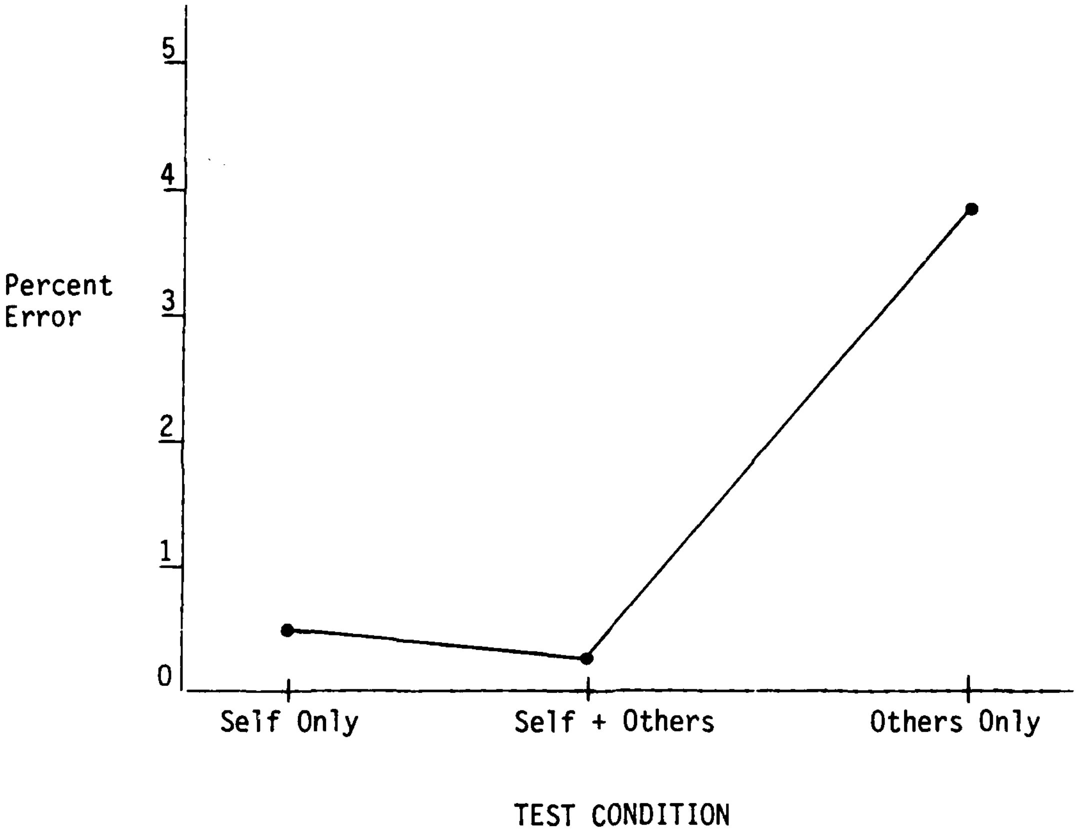


FIGURE 3-3.
NONRECOGNITION ERRORS BY TEST CONDITION

own speech pattern was not available and his voice inputs were compared to the speech patterns of only the other members of his group, nonrecognitions increased significantly.

3.4 Misrecognitions

As was done for nonrecognitions, an ANOVA was performed on the misrecognitions alone, to determine the effects of groups, trials and conditions. Table 3-5 presents the ANOVA summary table for misrecognitions.

A significant main effect of trials (F = 2.72, P < .05) is evident. The main effects of test condition and of groups were not significant, nor were any of the interaction effects. Mean misrecognition rates (in percent) are shown in Table 3-6, and the main effect of trials is portrayed graphically in Figure 3-4.

With regard to the main effect of trials, a Scheffé test for significance between pairs of means was performed to determine between which pairs of means the significant differences lie. Again, as was the case for total errors, the main effect of misrecognition errors by trials, reported in the ANOVA, could not be detected in the Scheffé or other appropriate range tests. Review of Figure 3-4 may clarify this finding. It can be seen that misrecognitions do vary somewhat as a function of trials. However, the greatest number of errors (Trial 1) was less than 1%, leaving little range for variability with a floor of zero. With the stringent per comparison alpha level imposed by the Scheffé test, the difference in range between trial one and trial five (where the least errors occurred) did not reach significance. All statistical results considered, the trials effect may best be viewed as a slight reduction of errors over trials, which may represent some practice effect. The authors, however, hesitate to consider this finding of any practical significance.

TABLE 3-5

ANALYSIS OF VARIANCE SUMMARY TABLE FOR MISRECOGNITIONS.

Source of Variance	df	MS	F
Group (G)	2	.16635	1.59
Error	12	.01483	
Test Condition (C)	2	.05295	1.42
C x G	4	.04837	1.30
Error	24	.03732	
Trials (T)	5	.02905	2.72*
T x G	10	.02065	1.93
Error	60	.01068	
C x T	10	.01363	1.09
C x T x G	20	.01184	.95
Error	120	.01249	

* p < .05

TABLE 3-6

MEAN MISRECOGNITION RATES (IN PERCENT)
FOR TEST CONDITIONS BY TRIALS.

TEST CONDITIONS

	Self Only	Self + Others	Others Only	$\overline{x}$ Trials
Trial 1	00.533	01.067	01.200	00.933
2	00.533	00.800	01.067	00.800
3	00.533	00.400	01.467	00.800
4	00.533	00.133	01.067	00.578
5	00.267	00.133	00.800	00.400
6	00.533	00.400	00.533	00.489
$\overline{x}$ Test Conditions	00.489	00.489	01.022	Grand $\overline{x}$ 00.667

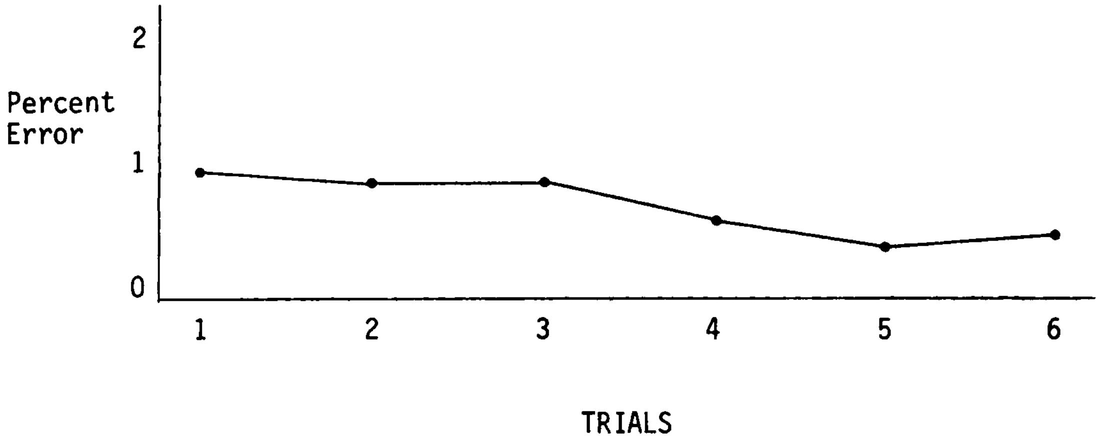


FIGURE 3-4
MISRECOGNITION ERROR RATES BY TRIALS.

4. DISCUSSION

Having presented the results of the present study, some implications of those results are now discussed.

4.1 Total Errors

There was no significant difference in total errors when subjects' speech input was tested against their own speech patterns versus testing against their own speech patterns plus the speech patterns of four other group members. However, when subjects tested against the speech patterns of the rest of the group, *without* access to their own speech patterns, total errors increased significantly. In positive terms, accuracy dropped from an average of 99.17% when subjects' own speech patterns were inaccessible and utterances were tested against the speech patterns of four other subjects. The statistical significance of the 4.01% reduction in accuracy simply means the change was unlikely to have occurred by chance. Whether or not 4.01% more errors is of practical significance depends on the type of errors made and the nature of their consequences.

In any event, the finding that the VRD is capable of recognizing, with greater than 95% accuracy, the speech of users for whom no speech patterns are available (speaker independence) is quite encouraging. Ninety-five percent accuracy in speaker independence opens the door for VRD's in a variety of settings. Speaker independence lends enormous freedom in situations where it is impractical for all potential users to train the VRD, or impossible for the VRD to retain more than a limited number of patterns even if all users could train it.

4.2 Nonrecognitions

Nonrecognitions accounted for over 93% of the variance in total errors across conditions. In effect, the increase in nonrecognitions in the speaker independence condition was responsible for the main effect of condition in total errors as well as nonrecognitions errors. Nonrecognitions averaged .34% in conditions where the VRD had access to the user's speech patterns and those of the other subjects (self only and self + others), but jumped to 3.82% when the user's own speech patterns were not available. This represents an increase of 3.48%. Still, even in the "speaker independent" mode; nonrecognitions only reduced accuracy to 96.18%. Figure 4-1 shows the relationship of type of errors by condition.

4.3 Misrecognitions

There were no significant effects of conditions on misrecognition errors. Misrecognitions remained around ½ to 1% under all conditions. In the self + others condition, this finding was not surprising, since the VRD usually chose the user's own speech pattern as the best candidate for a match with speech input. But in the others only condition, the current user's speech patterns were not present, forcing the VRD to base a decision on the speech patterns of other users. Under these circumstances the candidates for a match were poorer overall, yet the VRD showed no significant increase in misrecognitions. This is especially important since misrecognitions are the more problematic of the two types of errors, as explained earlier.

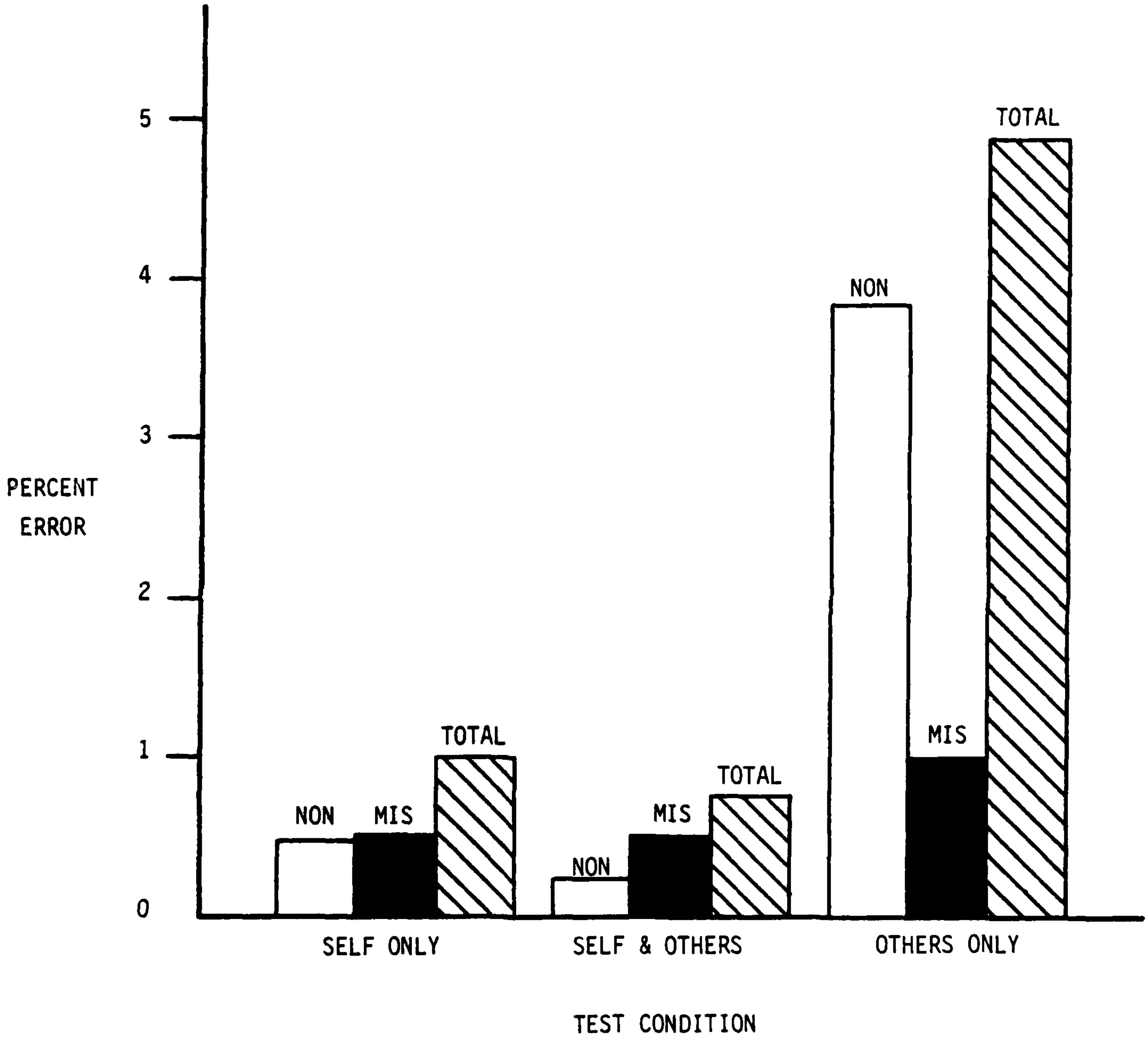


FIGURE 4-1
NONRECOGNITIONS, MISRECOGNITIONS, AND TOTAL ERRORS, AS A FUNCTION OF TEST CONDITIONS

5. CONCLUSIONS

The present research has shown that with current VRD technology, accuracy can be maintained at 99% when the speech patterns of four different users are combined with the current speaker's patterns in the VRD's memory.

Further, when the VRD has access to the speech patterns of four users, the speech input of an independent speaker (for whom the VRD has no speech patterns in memory) can be recognized with accuracy over 95%, and with no significant increase in misrecognitions. These results reflect favorably on the current capabilities of VRD technology and potential applications. Apparently, the algorithms employed (in the T600) allow enough variance on the appropriate dimensions to permit one person's speech input to correctly match the same utterance when spoken by a different person, while controlling variance or dimensions that would allow a speech input to incorrectly match a similar sounding utterance. The cost of this benefit is a fairly small increase in nonrecognitions (about 4%), the less problematic error.

While these findings are quite encouraging, some important issues should be noted. First, a note on the characteristics of the speakers is in order. Although no objective analysis was made of the voice patterns of the subjects, it is fair to say that most voices seemed to be about the same. With the exception of two subjects whose voices seemed to be slightly higher pitched and somewhat less clear ("raspier," perhaps), than others, no noteworthy differences were apparent in pitch, tone, or quality of the subjects' voices. Future research should attempt to quantify voice characteristics of subjects so that any relationship between performance on the VRD and specific voice characteristics can be elucidated. Second, in the others only condition the VRD accurately matched the speech of independent speakers for whom no speech patterns were available. However, <u>all</u> subjects had practice making voice inputs to a recognition criterion in which the training process was repeated for any utterance until the VRD accurately identified at least two out of three passes (see Methods). As a result of the training session, the subjects may have learned how to speak to the VRD for the best results. Therefore, the results of the current study may not generalize to naive speakers for whom the VRD has no speech record. The authors suggest the findings of the current study be taken in context, until further research can identify and quantify the significance of the initial training session.

6. REFERENCES

Brown, M.G., Engleman, L., Frame, J.W., Hill, M.A., Jenurich, R.I., and Toporek, J.D. *BMPD Statistical Software 1981,* Los Angeles: University of California Press, 1981.

Bruning, J.L. and Kintz, B.L. *Computational Handbook of Statistics (2nd ed.),* Glenview, Illinois: Scott, Foresman and Co., 1977.

Myers, J.L., Fundamentals of Experimental Design (2nd ed.), Allyn and Bacon, Inc., 1972.

Neter, J. and Wasserman, W. *Applied Linear Statistical Models,* Homewood, Illinois: Richard D. Irwin, Inc., 1974.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H. *Statistical Packages for the Social Services (2nd ed.),* New York: McGraw-Hill, 1975.

APPENDIX A

LIST OF UTTERANCES

WORD #	UTTERANCE	WORD #	UTTERANCE
0	ONE	25	SIERRA
1	YANKEE	26	APPLICATION
2	GARY POOCK	27	HUMAN FACTORS
3	CARRIAGE RETURN	28	CENTRAL EXPRESSWAY
4	IRAN	29	FILE TRANSFER PROTOCOL
5	SWEDEN	30	NINE
6	LOGIN POOCK	31	INDIA
7	ACCAT TITLE	32	LIMA
8	LOAD GLD3	33	POPPA
9	POOCK NPS PASSWORD	34	UNIFORM
10	THREE	35	KOREA
11	LOGOUT	36	INTERACTIVE
12	RED SPHERE	37	CONTINUOUS
13	SEVEN	38	CONTINUOUS SPEECH
14	MOVE IT DOWN	39	SYSTEM INTEGRATION
15	SPIROGRAPH	40	MIKE
16	CLOSE OUT CHARLIE	41	TANGO
17	UNITED STATES	42	WHISKEY
18	NORTH ATLANTIC MAP	43	ZULU
19	MEDITERRANEAN MAP	44	BANGLADESH
20	SIX	45	HOLLISTER
21	BRAVO	46	CORPORATION
22	DELTA	47	ADVANTAGES
23	FOXTROT	48	RADIOLOGY
24	ROMEO	49	AUTOMATIC RECOGNITION

DISTRIBUTION LIST

	No. of Copies
Library, Code 0142 Naval Postgraduate School Monterey, CA 93940	4
Dean of Research Code 012A Naval Postgraduate School Monterey, CA 93940	1
Library, Code 55 Naval Postgraduate School Monterey, CA 93940	1
COL Paul Cerjan 9th Infantry Division Fort Lewis, WA 98433	2
Professor Gary Poock, Code 55Pk Naval Postgraduate School Monterey, CA 93940	125